大数据对数字出版的应用价值初探

摘　要：大数据技术出现后，对数字出版产业的发展产生了深远的影响。无论在内容生产、流程管理、传播方式，还是交付形态等诸多方面，都表现出不同以往的特征，即生产主体多元化、流程管理协同化、传播方式场景化、交付形态多样化。数字出版逐渐摆脱传统出版的局限，向图书、期刊、报纸、广播电视以及互联网出版等多种业态融合的趋势发展。本文基于这一时代背景，从用户画像、产品发掘与过滤、出版流程协同、供需链结构变化以及知识服务的深化等方面阐述了大数据对数字出版的应用价值，认为通过大数据精准定位，实现用户体验的优化与迭代；构建知识关联体系，实现内容的发掘与过滤；整合企业内外资源，推动产品的协同运营；优化供需链结构，控制经营风险；加强资源深度开发，实现知识服务的更新迭代。

关键词：大数据；数字出版；知识关联；知识服务

中图分类号：G230　　**文献标识码：**A

文章编号：1671-0134（2019）09-016-06　　**DOI：**10.19483/j.cnki.11-4653/n.2019.09.001

文 / 赵宏源

前言

随着大数据的发展逐步迈入实质应用阶段，对社会各个层面都产生了重大的影响：行业层面上，决策依据更加全面和准确，有助于形成新的核心竞争力，对企业的生产模式、商业模式以及管理模式等都产生了深刻的变革，行业之间的界限模糊化；思维方式层面上，从满足于样本数据变为使用全体数据，从追求精确性到容忍混杂性，从关注因果关系到关注相关关系；智慧和战略层面上，数据的膨胀在“量”上引发混沌的同时，在“质”上可能开始显现出秩序，对社会发展和管理模式提出了新的挑战。[1]

在这个时代背景下，数字出版的发展逐步走向多元化：网络出版、移动出版、数据库出版、融合出版以及知识服务等。数字出版本身的特征也发生重大变化，从内容生产、管理过程、产品形态以及传播渠道的单纯的数字化[1]发展为内容生产主体的多元化、交付方式的多样化、服务过程的系统化以及内容推荐的智能化。内容生产主体的多元化体现为 UGC（用户生产内容）、PGC（专业生产内容）、OGC（职业生产内容）以及 DGC（设备驱动内容）等多种内容生产模式的涌现；交付方式的多样化如纸质图书、音频图书、VR/AR 图书、知识付费乃至知识服务等；服务过程的系统化体现在多层次的关联体系为读者提供完整的知识体系；内容推荐的智能化体现在用户画像基础上的精准推荐。

在数字出版上述特征的形成与演变过程中，大数据技术的应用是最主要的推动因素。随着数字出版产业的发展，大数据的应用价值越来越大。

1. 大数据对数字出版的重要价值

1.1 用户画像的构建与迭代

用户画像（persona）是美国交互设计师阿兰・库珀在《交互设计之路—让高科技产品回归人性》一书中提出的概念，persona 本意为面具，引申义为“面具”所表现出来的“角色”，主要作用在于描述用户特征。根据用户的现实特征和潜在特征，用户画像可以分为静态画像和动态画像两种。前者主要针对用户的统计学特征（如出生日期、性别、学历以及职业等）和行为特征（如衣食住行、娱乐、社交以及学习等消费习惯）；后者则针对用户的可能性、潜在性和趋势性特征。[2] 前者以用户需求为导向，重点在于应用场景中的用户行为模式，后者以生活愿景为导向的用户生活方式的描述，重点在于数据挖掘基础上的用户趋势预测。[3]

大数据视角下的用户概念，外延扩展到所有用户，既包括现实或者潜在用户，也包括跨行业潜在用户。界定用户的标准在于其价值提供和价值共创的结果，而非是否购买企业的产品 / 服务或者参与过企业活动。[4] 大数据基础上的用户画像通过用户标签化，构建数据模型研究“全样本”用户特征，有助于准确识别和描绘目标用户，有效联接用户的属性、行为与期待，极大地避免了主观因素的干扰；有助于及时掌握用户的异动，促进产品 / 服务的升级或者调整；有助于针对用户的个性化需求实现精准推荐。[5]

与其他行业比较，数字出版的用户的特点更加复杂，首先是用户角色的多重性。同一个用户既可能单一的服务使用角色，也可能是集内容的创作者、使用者、演绎

者以及传播者中某几个或者全部于一体的混合角色。其次是用户之间的关联性。基于对共同主题内容的关注和用户自身知识层次产生用户之间的碰撞效应或者同化效应，形成用户之间多种多样的关联关系。最后是用户特征的动态演变。随着用户知识量的增加，对内容的需求、内容产业链上扮演的角色以及社交关系等都处于持续的动态演变过程中。数字出版的目标用户的这些特点不仅要求准确描述用户的现实特征，进行静态画像，以便实现产品 / 服务的精准推送，更要着眼未来，准确把握用户需求的发展趋势，进行动态画像，以便捕捉用户兴趣与需求的变化，实现用户画像的完整构建与不断迭代。

1. 2 长尾产品的发掘与过滤

2004 年 10 月，《连线》杂志主编克里斯・安德森提出长尾理论，他认为，“只要存储和流通的渠道足够大，需求不旺或销量不佳的产品所共同占据的市场份额，可以和那些少数热销产品所占据的市场份额想匹敌甚至更大，即众多小市场可汇聚成与主流大市场相匹敌的市场能量”。长尾效应的意义在于“将所有非流行的市场累加起来会形成一个比流行市场还大的市场”[6]。

大数据时代的到来，人们获取的信息量越来越大，长期存在的信息不对称现象因此得到改善。与此同时，大数据也引发了选择困境，面对海量的信息人们无所适从，尤其对大量价值有限甚或垃圾信息难以甄别，从而形成了大数据实践应用中的悖论。反映在市场经济中体现为幂律分布[7]与长尾分布的并存，一方面，由于关注度的集聚效应，头部产品所占比例越来越小，单品销量越来越大；另一方面，长尾产品越来越多，质量参差不齐，人们的选择与甄别越发困难。这一问题的解决依然要回到大数据本身，通过完整的用户画像与产品特征的描述实现产品的精准推荐。

图书出版业属于小众行业，一方面，大多数图书的读者群有限，不可能有巨大的销量；另一方面，如此众多的图书充斥市场，结果是大多数图书很难找到目标读者，只有极少数的图书最终成为畅销书。[8] 进入数字出版阶段之后，这一特征体现的更加明显，存储和复制的边际成本趋近于零，创作门槛大幅降低，所出版的作品规模急剧扩张。幂律分布与长尾分布的两极化现象在数字出版业内更加剧烈，长尾产品的挖掘与过滤更加困难。通过大数据技术，有助于建立完整的用户画像，以识别用户需求，建立完整的知识关联体系，以导航定位发掘优质的长尾产品，通过用户评价以及作品相关的各种被引用关系等实现劣质长尾产品的过滤。

1. 3 出版流程的协同

所谓协同，就是指协调两个或者两个以上的不同资源或者个体，协同一致地完成某一目标的过程或能力。站在具有独立法人地位企业的角度，协同包括企业之间的协同和企业内部的协同。企业之间的协同包括纵向协同与横向协同，前者最典型的例子就是外包，后者即联合开发，双方或者多方组成项目组共同开发，获取的知识产权和利润按比例分摊。企业间协同根据紧密程度分为三个层次，即低耦合、中耦合和高耦合。企业内部的协同也分为三个层次，即产品结构的协同、专业技术的协同和内部各业务部门之间的协同。[9]

大数据背景下的企业协同，更注重资源整合意义上的协同。从内部而言，必须打破系统边界，打破部门壁垒，整合企业内外多源异构的数据，通过大数据分析，掌握企业自身优势与劣势、市场环境、用户特征与需求，改革并优化企业内部的管理与服务流程，为用户提供精准的产品 / 服务。从外部而言，企业不同阶段的发展需求必然不断产生各种或虚或实的团队，导致协同的边界不断扩大，相应的资源需求不断扩大而且多变。通过大数据分析，有效整合包括人力、组织、技术以及资金等在内的各种企业外部资源，并随时根据实际状况动态调整，实现企业内部之间以及企业之间信息的共享和利用，优化外部合作流程，从而大幅提升效率。[10]

相比其他行业，数字出版流程涉及的因素更为复杂，用户画像描述、选题策划、表达方式确定、交付方式选择、外部审校资源与作者资源整合、关联体系构建、交互功能设计与内容监管以及版权追踪监控等，都必须整合企业内部与外部的各种资源，针对不同需求进行相应的甄别与选择，尤其是面对快速多变的内容创作创新，更加要求协同的出版流程，实现产品 / 服务市场的高效运营。大数据技术的运用，有助于提升出版流程各个环节的市场化程度，据此实现内部部门之间的分工协调，外部资源的充分利用，进而提升数字出版企业的协同管理效率。

1. 4 供需链结构的优化

供需链的研究起源于 20 世纪 60 年代，一般是指由进行物料获取、加工物料成中间件或者成品，再将成品送到用户手中的一些企业或者企业部门构成的网络。参与供需链的主体包括供应商、生产商、配送中心、零售商以及最终用户。供需链涉及的内容包括参与供需链的有关实体之间的物流、信息流与资金流。[11] 根据供需链中起核心作用的企业类型，供需链模式可以分为以分销商为核心的供需链、以制造商为核心的供需链、以拍卖商为核心的供需链三种。[12] 有效的供需链管理有助于企业降低成本，提高竞争力；加速整个供需链上物流与信息流的流动，缩短产品流通周期；突出企业核心能力，加强竞争力。[13]

大数据技术的出现，通过对供需信息的量化与分析，降低供需链中的不确定风险，改变了传统的供需链模式，体现在两方面：一是供需从面对面模式（即传统的批量模式）到点对点（即根据消费需求定点生产与销售）模式的确立；二是供需分析预测的广度、深度和准确度的提高。第一个体现对于生产者而言意味着生产端与消费端沟通距离的缩短，上文中的三种供需链模式因之分化解体，个性化需求的满足逐步成为企业生产的主流模式；对于消费者而言意味着按需购买模式的盛行，用户根据商家基于大数据分析的推荐和自身实际需求选择合适的产品 / 服务。第二个体现传统的样本分析、概貌分析演变为精准分析，并以此为基础对未来趋势进行预测，根据趋势预测确定企业的战略决策。[14]

数字出版与包括传统出版在内的行业供需链相比，构成和运作模式都有很大不同。供需链的构成上有内容供需链、技术供需链以及载体供需链，其中，内容供需链从作者到出版单位到平台商（包括出版单位自建平台）到用户端；技术供需链从数据加工商（例如内容的数字化加工或者音视频加工）到出版单位再到技术开发商（平台或者数据库开发）；载体供需链从作者到出版单位到渠道商再到零售商最后到读者。无论每一条供需链除了中介性质的出版单位之外，都不是一般行业的一对一和一对多，而是多对多，而且供需链上的角色之间还存在重合性或者协作关系，前者如作者与读者角色的重合，后者如数据加工商与平台开发商之间的协作。大数据的应用，使整个供需链的上下游都发生改变，从相对单一的线性结构变为多元立体的网状结构[15]。通过对最终用户画像的描述，供需链上各个角色的分析，实现出版不同角色供需的精准匹配，优化供需链结构。

1.5 知识服务的深层次发展

知识服务，指以知识资源为对象，以信息知识的组织、关联、挖掘、重构等为手段，为用户提供个性化的智能服务。它是当今知识信息资源共享模式设计的核心指导思想，这种共享模式为用户提供的是针对特定问题的解决方案，即“知识”。[16] 知识包括显性知识和隐性知识。其中，显性知识指容易转化为符号的知识，可以语言表述从而传播；隐性知识隐藏在知识主体手中和头脑中，体现为技能、诀窍、洞察力、经验等无法用语言或者特定符号编码的知识。[17] 前者最典型的呈现方式就是出版物，尤其是图书，系统的论证与阐述、便捷的传播与交付，成为显性知识最佳的载体选择；后者最常见的表现例如师徒传承、实践练习以及头脑风暴等，“只可意会，不可言传”是其最主要的特征。两者同时又可以相互转化，通过社会化实现隐性知识之间的转化；通过外化实现隐性知识到显性知识的转化；通过整合实现显性知识之间的转化；通过内化实现显性知识到隐性知识的转化。[18] 通过显性知识与隐性知识相互之间不断的转化，实现知识本身的创新与传播。

大数据的出现使知识服务在知识生产与更新、典型特征、服务范围与模式等方面都产生了迥异于以往的变化。首先，知识生产主体多元化，既有领域内专家，也有普通用户，既有职业机构，也有个体生产，出现既有系统化的专业知识，也有碎片化的应用 / 普及知识，从而增强了知识关联体系构建的必要性。其次，在知识更新的方式上，既有质变性的系统化更新，也有知识交互过程中产生的量变性质的碎片化更新，知识交互同时具有了用户体验与知识输出的功能。再次，知识服务的典型特征体现在：面向自主需求和智慧服务、不确定性、强调用户参与、按需使用与付费以及基于群体创新等。[19]

从社会功能角度看，个人认为，数字出版与图书出版并无二致，那就是娱乐和教化。前者如网络文学、网络游戏、数字音乐以及网络动漫等；后者如电子图书、数字报纸、数字期刊、网络教育出版物、数据库出版物以及手机出版物等。从上文对大数据背景下数字出版和知识服务的分析可以看出，知识服务更强调用户为中心，根据用户的信息偏好以及具体需要，采取个性化的服务方式，为用户提供针对实际问题的解决方案等，因此，可以把知识服务视为教化领域数字出版的深化。[20] 依托大数据技术，知识服务既可以整合已有的存量系统化知识资源，又可以整合社会上的相关碎片化知识资源，还可以通过知识交互产生新的知识。在此基础上，构建完整的知识关联体系，实现系统知识和碎片知识的关联，自动建立新知识与已有知识的关联，消除“信息超载”“知识碎片化”和“信息孤岛”等问题，促进知识服务的深层次发展。[21]

2. 借力大数据，实现出版的转型发展

2.1 基于功能定位，实现用户体验的优化与迭代

所谓用户体验，简单讲就是用户的主观感受，主要指用户使用产品之前、期间和之后的认知印象和回应。[22] 影响用户体验的因素主要有四个，即产品策略、用户界面、技术以及运营等。[23] 大数据通过量化用户的愉悦度、参与度、接受度、留存度与任务完成度等指标，建立完整的用户静态与动态画像，改善影响用户体验的因素，实现用户体验的优化与迭代。

大数据在用户体验上的功能诉求可以分为三个，即基于销售的智能推荐、基于选题的需求分析和基于创作的能力分析。所以，用户体验的优化与迭代应从这三个方面展开。

首先，应对海量的用户数据进行最大限度的采集与整理，分析其静态特征，包括统计学特征和行为特征，从而找到产品的潜在用户，进而提供精准推荐服务，进行场景化的介绍，简化用户选择过程，提高用户购买体验。

其次，加强非结构化数据的采集整理，尤其是潜在细分领域中用户共同关注的话题、发表的言论、表达的思想等等，通过挖掘与分析，发现用户数据背后所反映的深层思想和预见趋势，进而策划细分市场选题，并且在选题策划过程中持续追踪这些变化，不断迭代，最终形成符合用户需求的选题与产品。

最后，通过隐性知识标引与隐性知识地图构建，寻找符合需求的潜在作者，并对作者的专业能力与创作能力进行分析。隐性知识标引通过隐性知识显性化的方式进行：其一是个人主动提供，包括个人详细信息、经历、创作的作品以及交互过程中的内容输出等，从中选取标识个人的标签；其二是他人评价，主要是在各种交互过程中周围人士对所输出的内容的各种评论、解读以及态度等，从外部视角做出更客观的评价。最后是系统自动发现，及时捕捉个人的隐性知识，例如人与人的互动过程中不自觉显示的某些技能等。[24] 隐性知识地图的构建主要通过分析与作者相关的数据之间的关联关系，包括作者与作品关系、领域内作者之间关系以及领域内作品之间关系等。根据作者的专业能力、创作能力与沟通能力确定作品的创作方式、表达方式以及传播协作等。

2.2 构建关联体系，实现内容的发掘与过滤

大数据的价值由大数据的关联性即大数据结构表达。[25] 大数据伴生的信息超载和信息孤岛现象与数字出版的幂律分布和长尾分布并存现象导致除少数头部内容外，大多数内容无法得到足够的展示机会为用户所知晓，部分劣质内容也无法及时处理，信噪比降低。解决这个问题，就必须回归大数据的价值根本，那就是通过关联体系的构建，充分运用知识之间存在的多类型、多维度、多层次的关联关系，实现知识的挖掘、发现、验证、甄别以及过滤等。[26]

运用大数据技术构建知识关联，实现优质内容的发现与劣质内容的过滤，可以从几个方面展开：

首先，通过基于大数据的自动标引、社会标引以及知识元标引等构建显性知识、知识内容以及学科知识等类型的已有知识关联，在此基础上，结合知识交互过程中的输出，不断将新产生知识纳入既有知识体系的同时，完善知识（包括交互过程中的输出知识，下同）的评价指标体系，进而建立隐性知识关联与个性化的主题知识关联，通过知识之间的相互印证和用户评价两个角度达到知识发现与过滤的目的。

其次，从知识与知识之间、知识与人之间以及人与人之间等三个维度完善关联关系，知识与知识之间的关联即上文显性知识之间的关联；知识与人之间的关联意味着必须对人即读者和作者进行标引，这种标引既包括统计学意义上的特征标引，也包括行为意义上的特征标引，在此基础上，通过知识与作者之间的多样化关联，确定知识内容的真伪。知识内容真伪的辨识通过两个途径实现：其一是类似文献分析中的共引和耦合分析，发现指定知识内容的关联关系，从而做出正确判断；其二是用户评价，通过多数人的看法得出正确结论。

最后是综合运用简单关联、时序关联和因果关联等规则，及时发现异常数据，清除垃圾账号与垃圾信息，减少噪声。通过简单关联规则中两类变量（例如交互过程中输出内容与原始内容）之间相关程度的测算，判断输出内容质量的优劣与输出动机；通过行为发生的时间对行为(例如内容输出的行为与时间)进行时序关联计算，判断输出者的勤勉程度或者输出内容的性质；通过关联的时序、强度、特异性、可重复性、一致性、合理性以及相似性等标准，确定因果关联成立的可能性，从而发现新知识、判断现有知识的正确与否。[27]

2.3 整合企业内外资源，推动产品协同运营

大数据对数字出版流程优化的作用主要通过数据分析与挖掘实现决策支持，为产品生命周期不同阶段的内外资源整合提供识别与选择、汲取与配置、激活和有机融合，构建企业价值链，[28] 促进开发过程中各个环节的有效衔接与资源利用率的最大化，及时发现并处理运营过程中的隐患，加快产品的更新迭代，实现产品的协同运营。

通过大数据技术整合资源，推动产品协同运营应从以下几个方面展开：

首先，在选题策划阶段应从内容和主创作者两个角度考虑。内容方面，通过大数据对现有海量文献数据分析与挖掘，根据关注度、词频统计、共引度以及耦合度等多种指标，及时发现学科研究热点方向、学科细分市场盲点（例如研究或者科普）、不同层次读者关注的重点等确定选题方向（即所谓思想），然后，通过用户画像分析用户的应用场景，据此确定不同应用场景的表达方式。主创作者方面，在通过现有海量文献数据分析与挖掘的同时，建立作者数据库，对作者的专业研究方向、学术关注度、媒体关注度、学术传播度、用户关注度、创作风格、粉丝群体以及动态行为等跟踪分析，根据作品应用场景精准选定作者。基于版权引进的选题分析则在此基础上增加版权作品输出区的市场分析、作者与作品在输入区的影响力与前景分析等。

其次，在内容生产阶段主要考虑两个方面：一是场景化的表达与迭代；二是内容关联的及时建立与更新。前者根据用户的应用场景细分化，并对不同场景的相关数据资源予以整合，实现内容的一次创作，多次发布。在此基础上，确定与应用场景相匹配的表达方式和交付方式，例如沉浸式深度或者浅度阅读、陪伴式阅读（亲子阅读可归入此类）、伴随式阅读（音频读物），可分别采用专业论述或者普及讲述，音频或者文字等。在创作过程中，不断根据用户反馈协助作者调整创作思路，完善交付设计，实现产品的更新迭代。后者则在内容发布更新的同时，不断建立完善内容关联，实现交付的增值与迭代，方式如上文所述。

最后，在产品传播阶段，通过大数据技术着重解决产品与应用场景的连接问题，提高产品交付效率。这个连接主要通过两条途径实现：一是通过大数据分析找到目标用户及其使用场景，通过用户画像实现产品的智能化精准推荐；二是通过针对性的诱因设计，刺激用户的主动传播行为，实现产品的裂变传播。其中，智能化推荐通过两种方式实现，一种是整合用户与应用场景连接渠道基础上的主动推荐，例如智能音箱最适合伴随式阅读场景，因此，通过智能音箱的实物分发渠道向用户推荐适合伴随场景的有声读物；另一种是基于关联（此处关联指普遍意义上的关联，不包括知识关联）规则挖掘的关联推荐，通过分析关联的提升度、置信度以及支持度等指标找到强关联规则，为用户提供扩展性推荐。

2.4 优化供需链结构，控制经营风险的同时降低运营成本

供需链优化的主要作用在于及时识别经营风险，降低运营成本。大数据对供需链优化的两个作用主要通过两端即上游供应端和下游需求端的历史数据分析研究，在发现与创造用户需求的前提下判断供应端的各种能力与风险，提高经营水平。

通过大数据技术优化供需链，实现经营效率的提升应从两方面展开：

从供应端角度重点加强能力分析，包括供应端的生产能力、响应能力、管理能力与风险防范能力等。数字出版的供应端包括内容供应者、技术供应者以及载体供应者，其中，内容供应者是重点，各种 UGC、PGC 与 OGC 都可以归入此类。首先通过大数据采集数字出版单位自身的业务数据、用户体验、交互数据等，分析并确定自身的定位、核心功能，进行动态修正与调整。在此基础上，通过对 PGC 与 OGC 等内容供应者的分析考察，建立基于知识生产的分享与协作网络，确定系统化的内容来源与供应节奏；建立 UGC 内容生产者的筛选与审核办法，并对内容建立严格的审查办法，确保输出内容的合法性；建立对 UGC 内容生产者的发掘与培养办法，逐步往 PGC 方向发展，提升其内容生产质量；结合人工审核与用户反馈，建立优质内容的筛选、评价与深加工流程。

从需求端角度重点在于通过对各种半结构化和非结构化数据的分析，主要是基于知识元分析的内容、内容交互过程中的输出内容（例如书评）以及各种自媒体内容等，发现空白或者尚未满足市场的内容需求，通过与之有关的相关关联与因果关联分析，确定所需基本的表达思想，进一步通过用户画像所呈现的应用场景确定内容的表达方式。其次是用户的基本特征数据、消费记录数据、地理位置数据、线上互动数据、线下行为数据、web 公共数据以及社交生成数据等[29]确定内容生产元素与用户数据之间的关系。最后通过与供应端内容生产者的匹配，确定不同内容生产者的分工合作，满足市场需求。

2.5 加强资源深度开发，实现知识服务的更新迭代

大数据对出版资源的深度开发在自有资源整合的基础上完成知识的组织、关联、挖掘与重构，完成知识服务基本架构的搭建，进一步拓展整合外部资源并及时更新，形成知识服务产品的更新迭代，帮助用户不断实现知识发现。

通过大数据整合资源，推动知识服务不断升级，主要通过以下几个方面展开：

首先是在根据上述目标用户画像，确定知识服务产品主题的前提下，组织所需的内容资源，对资源进行深加工，构建包括知识元标引、隐性知识标引、动态标引以及社会标引在内的知识标引体系，据此建立完整的知识关联。

其次是不断对知识服务产品运营过程中通过知识交互产生的输出内容进行标引，与已有知识内容建立关联；及时获取社会新产生的知识内容与信息资讯，在知识标引的基础上将其纳入知识体系。在形成动态知识关联的同时不断扩展知识体系的边界，实现知识服务的更新迭代。

最后是根据用户的行为特征与知识之间的连接关系建立个性化的关联关系，为用户提供个性化的知识体系，实现个人意义上的知识发现。例如，用户学习输入假说，动机可能是研究第二外语教学理论，也可能在研究知识服务运营，还可能在研究写作理论。因此，必须综合考察用户的浏览内容、创作内容（包含引文）、交互内容以及其他行为特征，据此为用户提供或二语教学，或知识服务运营，或写作理论研究的关联内容，提供精准的内容推荐，有效促进用户的知识吸收，推动知识输出，实现知识的分享与协作。

结语

与传统出版比较，数字出版融合图书、期刊、报纸、

广播电视以及互联网出版等多种形态，与知识服务的界限日趋模糊，不仅强调内容的系统性、关联性、交互性，更强调表达的场景化、内容的个性化和推荐的精准化，提高思想到达的有效性。这些特征的产生一方面是大数据在数字出版领域发展的必然结果；另一方面，通过大数据技术得到了强化。

大数据通过对各种信息数据资源的整合、分析、关联、挖掘与重构对用户画像、长尾经营、流程再造、供需链优化以及知识服务深化等方面产生了深刻影响，形成真正以用户为中心的内容生产、交互与交付格局，从根本上颠覆了传统出版的运营模式。在这种背景下，数字出版企业必须从用户体验优化、内容的发掘与过滤、产品的协同运营、供需链结构的优化以及知识服务的更新等方面入手，推动出版业的全面转型发展。媒

参考文献

[1] 姜浩端 . 大数据的本质及其可能的影响 [J]. 中国经济报告，2013（6）：16–22.

[2] 黄成群，余艳平 . 基于数字出版特征的现代出版人才技能探究 [J]. 出版参考，2017（8）：40–42.

[3] 代福平，辛向阳，张慧敏 . 用户动态画像 : 描述用户就是创造用户 [J]. 装饰，2018（3）：94–96.

[4] 张慧敏，辛向阳 . 构建动态用户画像的四个维度 [J]. 工业设计，2018（4）：59–61.

[5] 谢康，吴记，肖静华 . 基于大数据平台的用户画像与用户行为分析 [J]. 中国信息化，2018（3）：100–104.

[6] 郝胜宇，陈静仁 . 大数据时代用户画像助力企业实现精准化营销 [J]. 中国集体经济，2016（4）：61–62.

[7] 邓凤仪，邓海荣 . 大数据时代数字出版的“长尾效应”[J]. 出版发行研究，2014（10）：27–29.

[8] 胡海波，王林 . 幂律分布研究简史 [J]. 物理， 2005（12）：889–896.

[9] 练小川 . 幂律、长尾理论和图书出版 [J]. 陕西师范大学学报（哲学社会科学版）， 2007（3）：110–116.

[10] 汪伟 . 老生常谈论协同：协同的种类（一）[J].CAD/CAM 与制造业信息化，2013（2）：13.

[11] 刘古权 . 解读大数据与大协同 . 智协智造微信公众号 https://mp.weixin.qq.com/s/ECtg19d8IuhdBiXcLauAMg【EB/OL】. 2014（5）：15.

[12] 柴跃廷，韩坚，吴澄 . 敏捷供需链及其管理 [J].《中国机械工程》2000，11（3）：352–355.

[13] 李刚，刘鲁，陈安 . 供需链管理的运作模式研究 [J]. 高技术通讯，2001, 11（2）：71–73.

[14] 刘丽文 . 企业供需链管理的实施策略探讨 [J]. 计算机集成制造系统 –CIMS，2000（2）：15–18.

[15] 张晏魁 . 大数据与供应链物流服务发展关系浅谈 [J]. 经营管理者，2016（28）：325.

[16] 张弛，钟瑛 . 基于大数据的出版供需链 2+++ 构造探究 [J]. 科技与出版，2015（5）：105–107.

[17] 付蕾 . 知识元标引系统的设计与实现 [D]. 湖北：华中师范大学教育技术专业，2009：1–61.

[18] 张红太 . 浅谈隐性知识与显性知识的相互转化 [J]. 重庆科技学院学报（社会科学版），2011（7）：162–163.

[19] 赵士英，洪晓楠 . 显性知识与隐性知识的辩证关系 [J]. 自然辩证法研究，2001，17（10）：0–24.

[20] 秦晓珠，李晨晖，麦范金 . 大数据知识服务的内涵、典型特征及概念模型 [J]. 情报资料工作，2013（2）：18–22.

[21] 刘治 . 数字出版企业的知识服务研究 [D]. 湖南：湘潭大学公共管理学院，2011：1–57.

[22] 赵宏源 . 出版视域下的知识关联体系构建 [J]. 中国传媒科技，2019（1）：52–55.

[23] 李斌 . 重新定义用户体验：清空卖方立场 回归用户视角 [J].《中国商人》2018（7）：51–53.

[24] 老曹 . 影响用户体验的 4 个因素 . 人人都是产品经理 http：//www.woshipm.com/pd/1664.html【EB/OL】. 2012（5）：17.

[25] 秦铁辉，汪琼 . 试论专家型隐性知识地图的构建 [J]. 国家图书馆学刊，2007（2）：58–62.

[26] 刘文远，李少雄，王晓敏，等 . 大数据知识发现 [J]. 燕山大学学报，2014（5）：377–380.

[27] 赵宏源 . 出版视域下的知识关联体系构建 [J]. 中国传媒科技，2019（1）：52–55.

[28] 赵燕枫 . 似是而非的因果关联 [J]. 科学世界，2013（8）：82–83.

[29] 刘菲，严建渊 . 大数据时代基于价值链分析的企业流程再造案例研究 [J]. 物流工程与管理，2014（7）：226–227.

[30] 刘德群 . 大数据驱动下的影视文本创作和生产模式变革 [J]. 南京邮电大学学报，2019（4）：38–45.

（作者单位：上海世纪出版集团）